VoiceXML을 이용한 음성 인식시스템에서의 ASP 모듈 연구

장준식, 김민석, 윤재석*
*대진대학교 컴퓨터공학과

A Study On The ASP Module Using VoiceMXL in Automatic Speech Recognition System

Joonsik Jang, Minsuk Kim, Jaeseog Yoon*
*Department of Computer Engineering
Daejin University

요 약

본 연구에서는 VoiceXML의 용이한 확장성과 GSL(Grammar Specific Language)을 사용하여 사람이 말하는 자연어를 컴퓨터가 잘 이해할 수 있게 기호화 하고 이를 컴퓨터가 어떻게 인식하는가에 대해 다루어 보았다. 그리고 Voice Portal 항공정보시스템을 구축하여 사용자가 원하는 정보를 들려 줄 수 있게 하기 위한 ASP(Active Server Page)모듈을 작성하여 Voice Portal 항공정보시스템 상에서 그 효율성을 실험하여 보았다.

Abstract

In this research, it has been shown that how the computer can recognize and understand spoken natural language and its symbolization using VoiceXML and Grammar Specific Languge. In order for user to hear correct information, ASP Module has been revised and its effectivities has been experimented on the Voice portal airplane information system platform.

1. 서론

음성은 사람들 사이에 가장 기본적인 통신 수단이다. 컴퓨터와의 대화 수단을 보다 인간과 가깝게 하기 위해서는 음성인식(ASR : Automatic Speech Recognition)이 필요하게 되었으며 이러한 음성인식은 매우 다양한 응용분야를 가지며 음성과학과 컴퓨터기술의 발전에 의해 크게 발전되어오고 있다. 컴퓨터형태의 발전 추이가 저용량이면서 휴대 기능으로 변화함에 따라 과거 음성인식의 주를 이루어왔던 PC 환경 하 에서 뿐만 아니라, PDA, 이동통신 단말기 등의 모바일 환경에서 Web 컨텐츠에 접근 활용할 수 있는 최적화된 음성 인식 기술이 필요하게 되었다. 이러한 Web 관련 ASR연구가 최근 상당한 연구가 이루어졌으며 어떤 것들은 진행중인 것도 있다[1-5]. 이러한 기존의 마우스나 자판 입력을 대체하는 VUI(Voice User Interface)의 필요성과 이를 가능하게 해줄 새로운 컴퓨터언어인 VoiceXML (Voice eXtensible Markup language)의 필요성이 대두되었다. VoiceXML은 W3C에 의해 표준화된 인터넷과 음성을 바로 연결하여 연동을 할 수 있는 가장 효과적인 언어이다[6-7].

본 연구에서는 VoiceXML의 용이한 확장성과 GSL(Grammar Specific Language)을 사용하여 사람이 말하는 자연어를 컴퓨터가 잘 이해할 수 있게 기호화하고 이를 컴퓨터가 어떻게 인식하는가에 대해 다루어 보았다. 그리고 사용자에게 어떤 컨텐츠를 제공하고자 할 때 그 내용이 사용자가 어떤 것을 원하느냐에 따라서 다른 내용을 제공해야 한다면 그 많은 내용을 VoiceXML만으로 수백 수천에 이르는 페이지를 작성하여서 들려준다면 정말 비효율적일 것이다. 그래서 본 연구에서는 Voice Portal 항공정보시스템을 구축하여 사용자가 원하는 정보(예를 들면 원하는 날짜의 비행정보)를 들려 줄 수 있게 하기 위한 ASP(Active Server Page)모듈을 - 원하는 날짜의 비행 정보를 말해 주는 것을 ASP를 이용하여 사용자가 말한 내용을 데이터베이스에서 검색하여 그 내용을 VoiceXML로 변환하여 사용자에게 들려주는 모듈 - 작성하여 Voice Portal 항공정보시스템 상에서 그 효율성을 실험하여 보았다.

2. 웹기반 음성인식 시스템의 구축

2.1 웹 기반 음성 인식 시스템 개요

그림 1은 본 연구에서 사용한 음성 포탈 서비스를 위한 시스템을 나타낸 것이다. 그림 1에서 VoiceXML Gateway는 VoIP(Voice Over IP), 전화, 모바일, 핸드폰 등과 같은 음성 입력을 받을 수 있는 모든 디바이스와의 인터페이스를 담당한다. 웹서버에는 각종 음성 파일과 grammar, 스크립트, 멀티미디어 자원 등이 탑재되어 있고 이것을 VoiceXML Gateway를 통해 사용자에게 정보를 제공한다. 이 웹서버내의 컨텐츠는 VoiceXML형 문서나 HTML형 문서로 변환되어 VoiceXML Gateway나 일반 웹 브라우저로 전송하게 된다. XML과 같은 메타 언어를 사용하게 되면 웹 문서를 원하는 형식으로 변환이 가능하다. 이 점을 이용하여 웹서버 내의 문서를 HTML이나 VXML로 변환하는 과정만 더하게 되면 컨텐츠들을 통합 할 수 있다. 본 연구에서는 이것을 구체적으로 구현하여 가상의 항공사 사이트에서 음성으로 비행 정보를 제공하도록 하였다.

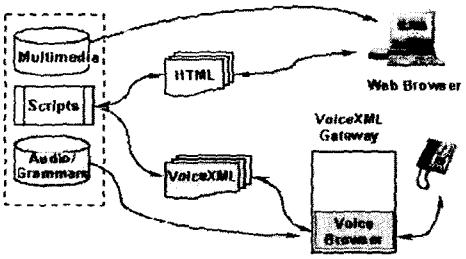

그림 1. 웹 기반 음성 시스템 개요도

2.2 실험 환경

본 연구의 실험 환경은 Pentium III 600MHz CPU, 256MB의 윈도우즈 2000 Professional을 웹 서버 및 데이터 베이스 서버로 사용하였다. 데이터베이스는 MS Access 2000을 사용하였으며 AMD Athlon 900MHz CPU와 512MB의 메모리를 탑재한 윈도우즈 2000 advanced 서버시스템을 Voice Web Server로 사용하여 하드웨어 플랫폼을 구축하였다. VWS(Voice Web Server)에는 음성인식엔진으로 Nuance사의 NSR(Nuance Speech Recognition) 7.04버전을 사용하였고, 웹서버에서 전송되는 VoiceXML문서의 interpreter로 Nuance VWS 1.2 Beta를 사용하기 위해 설치하였다. 이 것은 그림 1에서 보인 VoiceXML Gateway 역할을 한다. 기타 실험에 사용한 프로그램으로는 자바 실행 환경을 위해 JDK(Java Development Kit)1.3.1버전과 VWS 제어를 위해서 jakarta tomcat 3.2.3 버전을 설치하였다[8].

본 연구의 실험에 사용된 웹 기반 비행 음성정보 시스템의 구축도는 그림 2에 나타냈다. 사용자가 음성 입력이 가능한 입력 디바이스를 이용하여 서비스를 요청하게 되면 Voice Web Server는 음성인식엔진을 통해 요청을 받아들이고 사용자에게 전달할 컨텐츠를 웹서버에 요청하게 된다. 웹서버에서는 일반 VXML 문서일 경우에는 바로 Voice Web Server로 전송하게 된다. 본 연구에서는 VXML의 방대한 내용을 ASP 모듈을 이용하여 간결하게 줄였는데 이 때에서처럼 ASP문서를 요청하면 IIS(Internet Information Service)에서 ASP문서를 컴파일 후에 VXML 문서로 변환하여 전송하게 된다. ASP는 Database에 접근하여 요청하는 내용을 검색, 삽입, 삭제, 갱신을 하여 그 내용을 VXML로 변환한다. 사용자가 요청하는 내용에 따라 각각 다른 컨텐츠를 가진 VXML로 전송을 하게 되므로 상황에 따라 컨텐츠를 생성한다.

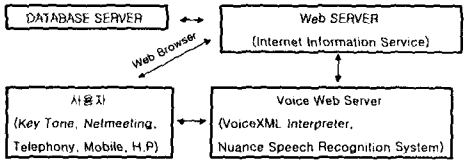

그림 2. 웹 기반 비행 음성 정보 서비스 실험 구축도

3. ASP를 이용한 모듈화

그림 3은 "welcome to service."라는 음성 정보를 사용자에게 들려주기 위해 VXML로 작성한 문서이다. 그림 4는 ASP 스크립트 언어를 사용하여 위와 같은 기능을 하도록 작성한 것이다. 그림 4의 문서가 IIS내에서 컴파일 되면 그림 3과 같은 형태로 변환된다[9].

그림 4에서 '<!-- #include file="vxml.asp" -->' 부분은 본 실험에서 작성한 VXML coding module이 포함된 파일을 포함하는 것이다. vxml.asp란 파일 내에는 위의 ASP 문서를 VXML로 변환하기 위한 함수들이 들어 있다. 원하는 VXML문서를 간편하게 변환하기 위해서는 이 함수들을 호출하여야 한다. VXML module을 포함시킴으로써 VXML문서를 간단한 함수 호출로 만들어 낼 수 있다. 함수를 사용하지 않는다면 코딩은 보다 복잡해 질 것이고 이해도 어려워질 것이다. 본 논문에서는 동적인 컨텐츠(비행 정보)를 만들어 낼 수 있도록 ASP모듈을 설계하여 그 효율성을 웹 기반 비행 음성 정보 시스템에서 실험하였다.

```
<?xml version="1.0"?>
<!DOCTYPE vxml PUBLIC '-//Nuance/DTD VoiceXML 1.0b//EN'
'http://community.voxeo.com/vxml/nuancevoicexml.dtd' >
<vxml version="1.0">
  <form id="welcome">
    <block>
      welcome to service.
    </block>
  </form>
</vxml>
```

그림 3. VoiceXML의 기본 예

```
<!-- #include file="vxml.asp" -->
<%
Response.ContentType = "text/xml"
call Header
call VxmlStart(NULL, NULL, NULL, NULL)
  call FormStart("welcome", NULL)
    call BlockStart(NULL, NULL, NULL)
      call Prompt("welcom to service", NULL, NULL, NULL, NULL)
    call BlockEnd
  call FormEnd
call VxmlEnd
%>
```

그림 4. ASP 모듈을 사용한 예

3.1 ASP 모듈 설계 및 실험

ASP의 기본모듈을 이용하여 비행 정보 컨텐츠를 생성하는 ASP 모듈을 설계하여 Voice Portal 음성인식시스템에서 실험하였다. 사용한 비행정보 데이터베이스는 비행번호, 출발지, 도착지, 출발월, 출발일, 출발시간, 도착월, 도착일, 도착시간, 여석, 대기자 인원의 속성을 가진다. 여기서 사용된 음성은 일반 사람의 음성을 조합하여 재생하도록 하였다.

본 실험에서 설계된 ASP 모듈은 그림 5에서 보여 주는 것처럼, 비행 정보를 생성하는 함수의 이름를 FlyingInfoPlay라 했다. 여기에 사용된 인수로는 service_type, flying_num, origin, destination, month, day, hour, period가 있다. FlyingInfoPlay 함수는 서비스의 타입(비행정보, 예약, 예약확인 등)이나 비행번호의 유무, 조회할 기간에 따라 각각의 다른 VXML 컨텐츠를 생성한다. 비행번호가 있을 시에는 이 고유의 비행 번호를 가지고 데이터베이스에서 검색하여 그 내용을 가져와서 컨텐츠를 형성하지만 이것이 없을 경우에는 출발지나 도착지, 출발월·일·시간 등으로 검색한다. 기간이 당일이나 주말 중에 선택을 할 경우 그것에 맞는 기간 동안의 비행 정보를 생성한다.

```
<!--#include file=" /vxml.asp"-->
<?xml version="1.0"?>
<!DOCTYPE vxml PUBLIC '-//Nuance/DTD VoiceXML 1.0b//EN'
                    'http://community.voxeo.com/vxml/nuancevoicexml.dtd' >
<vxml version="1.0">
  <form id="welcome">
    <block>
      <%
      Call FlyingInfoPlay("flying_information", Null, request("origin"), request("destination"),
          month(now), day(now), hour(now),request("to_we"))
      %>
    </block>
  </form>
</vxml>
```

그림 5. 비행 정보 컨텐츠 생성을 위해 설계된 ASP 모듈

그림 5의 문서가 웹서버의 IIS내에서 컴파일 되면 그림 6과 같은 VXML 문서가 생성되는데 그림에서는 지면관계상 생성된 VXML 파일의 일부분만 보여주었다. FlyingInfoPlay 함수에 다른 인수가 입력될 경우에 이 문서의 컨텐츠는 다른 내용을 담게 된다.

그림에서 보듯이 ASP 모듈을 사용하여 모듈화 하게 되면 훨씬 더 내용이 함축되어 있으며 동적으로 컨텐츠를 생성해 낼 수 있음을 알 수 있다. 본 연구에서 구축한 웹 기반 음성인식 시스템에서 실험한 결과 데이터베이스 컨텐츠를 음성신호로 변환하여 훌륭하게 전달하는 것을 확인하였다.

그러나 위의 ASP 모듈은 그 양이 많은 관계로 각 페이지를 로딩할 때마다 전체 ASP 모듈을 포함하게 된다. 이것은 로딩 시간이 늘어나는 문제점이 있다. 다시 말해 쓰이지 않는 모듈도 모두 포함하게 되는 문제점이 발생한다. ASP 문서 내에서 <% ...%>내에 작성된 내용만 ASP 컴파일러에서 컴파일 하게 하고 다른 내용은 그냥 그 자체로서 전송하게 하면 앞의 문제점들을 해결할 수 있었다. 따라서 본 실험에서는 문서 내에 동적 컨텐츠를 생성할 필요가 있는 부분만 ASP 모듈로 유효하도록 설계, 코딩하였다.

```
<?xml version="1.0" encoding="EUC-KR" ?>
- <vxml application="../main.vxml" version="1.0">
  - <form id="start">
    - <block name="block1">
      - <prompt>
          <audio src="/prompts/city/seoul.wav" />
          <audio src="/prompts/word/from.wav" />
          <audio src="/prompts/city/busan.wav" />
          <audio src="/prompts/word/to.wav" />
          <audio src="../../prompts/month/9.wav" />
          <audio src="../../prompts/month/month.wav" />
          <audio src="../../prompts/day/20.wav" />
          <audio src="../../prompts/day/4.wav" />
          <audio src="../../prompts/day/day.wav" />
          <audio src="../../prompts/hour/pm.wav" />
          <audio src="../../prompts/hour/10.wav" />
          <audio src="../../prompts/hour/1.wav" />
          <audio src="../../prompts/hour/hour.wav" />
          <audio src="../../prompts/minute/30.wav" />
          <audio src="../../prompts/minute/3.wav" />
          <audio src="../../prompts/minute/minute.wav" />
          <audio src="../../prompts/month/9.wav" />
          <audio src="../../prompts/month/month.wav" />
          <audio src="../../prompts/day/10.wav" />
          <audio src="../../prompts/day/5.wav" />
          <audio src="../../prompts/day/day.wav" />
          <audio src="../../prompts/hour/pm.wav" />
          <audio src="../../prompts/hour/7.wav" />
          <audio src="../../prompts/hour/hour.wav" />
          <audio src="../../prompts/minute-1/40.wav" />
          <audio src="../../prompts/minute-1/2.wav" />
          <audio src="../../prompts/minute-1/minute-1.wav" />
          <audio src="../../prompts/seat/50.wav" />
          <audio src="../../prompts/seat/4.wav" />
          <audio src="../../prompts/seat/seat.wav" />
```

그림 6. 설계된 ASP 모듈에 의해 생성된 VXML 파일

3.2 GSL(Grammar Specification Language)

Grammar는 사용자가 수행하고자 하는 것이나

제공할 정보를 문자열로 반환하게 하게 하는 것으로 VXML 표준안에는 그 포맷을 명세 하거나 특별한 포맷의 지원을 필요로 하도록 하지 않았다. 따라서 이것은 음성 인식 엔진 자체적으로 포맷을 가지며 반환한 값만을 VXML에서 처리한다. 본 실험에서는 이 grammar 포맷중의 하나인 Nuance사에서 제공하는 GSL(Grammar Specification Language)을 사용하였다.

GSL은 VUI(Voice User Interface)를 구현하는데 아주 중요한 구실을 하게 된다. 사용자는 그냥 "예약"이라고 하는 사용자도 있을 것이고 "예약하고 싶어요", "비행예약" 등과 같이 말할 수도 있다. 이중에서 예약이란 단어는 꼭 들어가게 된다. 다시 말해 사용자가 수행하고자 하는 것이나 입력하는 내용을 VXML에 전달하게 되고 VWS는 이것을 기반으로 처리한다. 이것은 일반 웹 프로그래머들이 회원 가입·인증의 폼의 텍스트 필드에 값을 넣는 것과 비슷한 원리이다. 입력 값으로 음성과 문자열 중에서 어느 것을 선택하느냐의 차이라고 할 수 있다. 음성으로 입력된 값이 VXML의 field 태그의 속성인 name에 지정된 변수에 저장된다. 이것과 비슷하게 html에서는 이것을 <input type="text" name="memberid">로 표현할 수 있다. 이 필드의 값이 모여 폼을 형성한다. grammar는 대부분 비슷한 형식을 취하고 있는데, 그림 7은 본 실험에서 GSL을 사용하여 작성한 한 예를 나타낸 것이다. 이것은 예약 체크나 취소 예약 방법을 택하게 하는 grammar로써 사용자가 말한 내용에 따라 각각 다른 출력을 하도록 작성한 것이다. 비행 번호로 예약을 원하면 출력을 number로 시간이나 지역으로 예약을 원하면 nomal을 출력하고, 확인은 check, 취소는 cancel을 반환하도록 하였다.

```
Reservation [↵
    ( ?( i want to ) ?flying number) { <choice number>}↵
    |↵
    (?( i want to ) area) (?( i want to ) time) (?( i want to ) nomal)↵
    |↵
  { <choice nomal>}↵
    (?( i want to ) ?reservation check) { <choice check>}↵
    (?( i want to ) ?reservation cancel) { <choice cancel>}↵
  ]↵
```

그림 7. 예약 방법 및 확인 체크를 위한 GSL의 예

4. 결론

본 연구는 VoiceXML이란 표준화된 마크업 언어 사용하여 웹 기반의 음성 인식 시스템을 구축하여 전화나 기타 음성 입력이 가능한 디바이스를 사용하여 웹 브라우징을 할 수 있음을 보였고, 기존의 웹 프로그래밍 언어의 하나인 ASP를 사용하여 좀 더 복잡한 구조를 가진 컨텐츠를 생성하고 이를 모듈화하여 재사용이 가능하게 할 수 있음을 보였다.

이제 웹은 더 이상 눈으로 보여지는 것뿐만 아니라 음성으로 컨텐츠를 전달하는 것이 가능하다. VXML과 기존의 프로그래밍 언어를 사용하면 다이나믹한 컨텐츠를 생성 가능하며 이로써 다양한 서비스를 구현할 수 있음을 알 수 있었다. 더불어 VXML이 XML을 기반으로 만들어진 마크업 언어인 만큼 얼마든지 확장도 가능한 것이다. 이러한 사실이 기존의 웹사이트와 음성 정보 서비스의 컨텐츠를 통합함으로써 웹사이트의 내용을 바로 음성으로 브라우징하고 그 컨텐츠를 음성으로도 전달할 수 있는 가능성을 열었다하겠다.

참고문헌

[1] K. Kondo and C.Hemphill. A WWW browser using speech recognition and its evaluation.
Systems and Computers in Japan, pages 57-67, Sep. 1998.

[2] V. Digalakis, L. Neumeyer, and M. Perakakis. Quantization of cepstral parameters for
speech recognition over the World Wide Web. IEEE Journal on Sel Areas in Communications, pages 82-90, Jan. 1999.

[3] Z. Tu and P. Loizou. Speech recognition over the Internet using Java. IEEE International
Confrence on Acoustics, Speech, and Signal Processing, Phoeix, AZ: pages 2367-70, Mar.
1999

[4] C.T.Hemphill and Y.K.Muthusamy. Deve loping Web-based speech applications. European
Conference on Speech Communication and Technology, pages 895-898, Sep. 1997.

[5] D. Vaufreydaz, J. Rouillard, and M. Akbar. A network architecture for building
applications that use speech recognition and/or synthesis. European Conference on
Speech Communition and Technology, pages 2159-62, Sep. 1999.

[6] http://www.w3.org/voice

[7] http://www.voicexml.org.

[8] http://www.nuance.com

[9] http://community.voxeo.com